# Productivity and Efficiency at Bank Holding Companies in the U.S.: A Time-varying Heterogeneity Approach

Guohua Feng*
Department of Economics
University of North Texas
Denton, Texas 76203
U.S.A.

Bin Peng
Department of Economics
University of Bath
Bath, BA2 7AY
U.K.

and

Xiaohui Zhang
Department of Economics
University of Exeter
Exeter, Devon, EX4 4PU
U.K.

July 10, 2017

*Corresponding author. Email: guohua.feng@unt.edu; Phone number: +1 (940) 565 2220.

## Abstract

This paper investigates the productivity and efficiency of large bank holding companies (BHCs) in the United States over the period 2004–2013, by estimating a translog stochastic distance frontier (SDF) model with time-varying heterogeneity. The main feature of this model is that a multi-factor structure is used to disentangle time-varying unobserved heterogeneity from inefficiency. Our empirical results strongly suggest that unobserved heterogeneity is not only present in the U.S. banking industry, but also varies over time. Our results from the translog SDF model with time-varying heterogeneity show that the majority of large BHCs in the U.S. exhibit increasing returns to scale, a small percentage exhibit constant returns to scale, and an even smaller percentage exhibit decreasing returns to scale. Our results also show that on average the BHCs have experienced small positive or even negative technical change and productivity growth.



*JEL classification*: C11; D24; G21.

*Keywords*: Productivity and Efficiency; Bank Holding Companies; Translog Stochastic Distance Frontier Model with Time-varying Heterogeneity.

## 1. Introduction

The productivity and efficiency of the U.S. banking industry has received considerable attention in the past three decades, reflecting the transformation of the industry caused by regulatory changes and technological and financial innovations (Wheelock and Wilson, 2012; Hughes and Mester, 2013; Davies and Tracey, 2014). One line of research in this area that has recently attracted increasing interest focuses on how to deal with the problem of unobserved heterogeneity among banks — an important issue given the widespread unobserved heterogeneity among U.S. banks (Cole *et al*., 2004; Berger *et al*., 2005; Rossi, 1998; Rosen, 2003). Studies along this line of research include Mester (1997) and El-Gamal and Inanoglu (2005). This study aims to contribute to this line of research by applying a stochastic distance frontier model, which allows for time-varying unobserved heterogeneity, to the analysis of the productivity and efficiency of bank holding companies (BHCs) in the U.S.[1]

There is evidence that time-varying unobserved heterogeneity is widely present in the U.S. banking industry. Consider, for example, bank asset quality. Studies (Hughes and Mester, 1993, 1998) suggest that the quality of a bank's assets can influence the bank's costs in a variety of ways (Hughes and Mester, 1993; Hannan and Hanweck, 1988). Thus, it would be desirable to incorporate a vector of variables characterizing bank asset quality when modeling the bank's production process. However, this is technically difficult for the following two reasons. First, asset quality is hard to measure (i.e., "unobserved"). Second, the effects of asset quality are "time-varying" due to changes in bank regulations, economic conditions, loan approval processes, and so on. To give another example, previous studies (Demsetz and Strahan, 1997; Dick, 2006; Hirtle, 2007) suggest that the geographic reach and local branch density of a bank's branch network has an important impact on the bank's cost structure, and thus should be accounted for when assessing productivity and efficiency. However, as with the case of asset quality, it is difficult to quantify this time-varying bank characteristic, "due to the lack of detailed branch data across a large number of institutions"

[1] There are two reasons why our analysis focuses on BHCs rather than individual commercial banks. First, total assets controlled by BHCs accounts for 99 percent of the industry assets in 2012 (Federal Reserve Board Annual Report, 2012). Second, important business decisions are typically made at bank holding company level (Stiroh, 2000).

(Hirtle, 2007). Besides these two examples, there are many other bank characteristics that are hard to measure and time-varying, such as "too-big-to-fail" factors (Davies and Tracey, 2014) and level of risk-taking (Hughes and Mester, 2013). These examples suggest that it is important to account for time-varying unobserved heterogeneity when investigating the productivity and efficiency at BHCs in the U.S.

The purpose of this paper is to apply a new stochastic distance frontier (SDF) model, which allows for time-varying unobserved heterogeneity, to BHCs in the U.S. The model, which we call the "translog SDF model with time-varying heterogeneity", is obtained by adding multiple time-varying individual effects to the standard translog SDF model. Following Ahn *et al.* (2013), the time-varying individual effects are modeled by a "multi-factor structure". Formally, the multi-factor structure is written as: $\boldsymbol{f}_t'\boldsymbol{\gamma}_i$, where $t = 1, 2, ..., T$ indexes time; $i = 1, \cdots, K$ indexes firms; $\boldsymbol{\gamma}_i$ is a vector of unobserved firm-specific variables; and $\boldsymbol{f}_t$ is the corresponding vector of time-varying parameters. The model thus has three error terms, with one capturing statistical noise, a second capturing inefficiency, and a third capturing time-varying unobserved heterogeneity.

It is worth noting that the multi-factor structure is also used in Anh *et al.* (2007), but for a different purpose. Specifically, in Anh *et al.* (2007) the multi-factor structure is employed to model firm-specific time-varying technical inefficiency, whereas in this paper it is used to capture time-varying unobserved heterogeneity. In addition, $\boldsymbol{f}_t$ and $\boldsymbol{\gamma}_i$ also have different interpretations in the two papers. In Anh *et al.* (2007), $\boldsymbol{f}_t$ represents common drivers of technical inefficiency and $\boldsymbol{\gamma}_i$ represents firm-specific responses to the shocks. In contrast, in this paper $\boldsymbol{\gamma}_i$ represents unobserved firm-specific variables and $\boldsymbol{f}_t$ measures the time-varying effects of $\boldsymbol{\gamma}_i$ on the dependent variable.

In assessing the productivity and efficiency of BHCs in the U.S., we use the output-distance-function-based productivity index proposed by Orea (2002) and Lovell (2003). This index has two desirable properties: 1) it is valid under different market structures and returns to scale; and 2) it allows for scale effects. These advantages imply that the productivity measure used in this study is theoretically correct even when we do not have a priori information about the market structures and returns to scale of the U.S. banking industry.

We choose to use a Bayesian approach to estimate the SDF model with time-varying heterogeneity. There are two reasons for this choice. First, we can easily obtain exact inferences on productivity and efficiency measures through the posterior distributions of these measures. Second, it is computationally attractive, because it only involves relatively straightforward modifications to the Bayesian formulations of standard stochastic frontier models (Koop and Steel, 2003; O'Donnell and Coelli, 2005).

Finally, we apply the translog SDF model with time-varying heterogeneity to BHCs in the U.S. over the period 2004–2013. Our empirical results provide strong evidence that the SDF model with time-varying heterogeneity is preferred to both the standard translog SDF model and the true random effects model. The superiority of the former model suggests that unobserved heterogeneity not only exists in the U.S. banking industry, but also varies over time.

The rest of the paper is organized as follows. In Section 2, we briefly discuss the output-distance-function-based productivity index, its three components, and returns to scale. In Section 3, we present the SDF model with time-varying heterogeneity. In Section 4, we discuss the Bayesian procedure for estimating the model. Section 5 describes the data. In Section 6, we apply our methodology to BHCs in the U.S., discuss the effects of incorporating time-varying unobserved heterogeneity, and report our estimates of total factor productivity growth and its components. Section 7 concludes the paper.

## 2. The Output-Distance-Function-Based Productivity and Efficiency Measures

### *2.1. Output distance functions*

We start by defining the output distance function. Consider the case of a multi-input multi-output production technology, where a bank holding company (BHC) uses the $N \times 1$ input vector $\boldsymbol{x}^t = (x_1^t, x_2^t, ..., x_N^t)'$ to produce the $M \times 1$ output vector $\boldsymbol{y}^t = (y_1^t, y_2^t, ..., y_M^t)'$ at time $t = 1, 2, ..., T$. Following Färe and Primont (1995), the production technology can be described by the technology set

$$P^t\left(\boldsymbol{x}^t\right) = \left\{\boldsymbol{y}^t : \boldsymbol{y} \text{ is producible from } \boldsymbol{x}\right\}.$$

The production technology satisfies a standard set of axioms including convexity, strong disposability, closedness and boundedness.

Färe and Primont (1995) show that this technology can also be described using an output distance function

$$D_o^t\left(\boldsymbol{y}^t, \boldsymbol{x}^t\right) = \inf_{\theta}\left\{\theta > 0 : \frac{\boldsymbol{y}^t}{\theta} \in P^t\left(\boldsymbol{x}^t\right)\right\}. \tag{1}$$

It gives the minimum amount by which an output vector can be deflated and remains producible with a given input vector. The output distance function is non-decreasing, convex and linearly homogeneous in outputs, and non-increasing and quasi-convex in inputs — see Färe and Grosskopf (1994, p. 38).

Following the common practice of modeling the effect of time through an exogenous time variable, $t$, the output distance function in (1) can be rewritten as $D_o(\boldsymbol{x}, \boldsymbol{y}, t)$. As indicated by (1), $D_o(\boldsymbol{x},\boldsymbol{y}, t) \leq 1$. Deviation of the output distance function from one, due to technical inefficiency, can be accommodated as follows,

$$D_o(\boldsymbol{x}, \boldsymbol{y}, t)\psi(t) = 1, \tag{2}$$

where $\psi(t) \geq 1$.

*2.2. The output-distance-function-based productivity and efficiency measures*

Following Lovell (2003) and Orea (2002), we use the following output-distance-function-based productivity index to assess the productivity and efficiency of BHCs in the U.S.

$$\frac{d \ln TFP}{dt} = \sum_{m=1}^{M} \tilde{\omega}_m \dot{y}_m - \sum_{n=1}^{N} \omega_n \dot{x}_n, \tag{3}$$

where

$$\tilde{\omega}_m = \frac{\partial \ln D_o\left(\boldsymbol{y}, \boldsymbol{x}, t\right)}{\partial \ln y_m},$$

and

$$\omega_n = \frac{\partial \ln D_o\left(\boldsymbol{y}, \boldsymbol{x}, t\right) / \partial \ln x_n}{\sum_{k=1}^{N} \partial \ln D_o\left(\boldsymbol{y}, \boldsymbol{x}, t\right) / \partial \ln x_k}.$$

It is straightforward to show that the following restrictions on $\tilde{\omega}_m$ and $\omega_n$ hold

$$\sum_{m=1}^{M} \tilde{\omega}_m = 1 \text{ and } \sum_{n=1}^{N} \omega_n = 1,$$

where the former holds by the linear homogeneity of the output distance function in outputs and the latter by definition.

We now define the-output-distance-function measure of returns to scale. Let $\varepsilon_n$ denote the elasticity of the output distance function with respect to $x_n$, i.e., $\varepsilon_n = \partial \ln D_o\left(\boldsymbol{y}, \boldsymbol{x}, t\right) / \partial \ln x_n$. The output-distance-function-based measure of returns to scale ($\varepsilon$) can then be defined as in Caves *et al.* (1982)

$$\varepsilon = -\sum_{n=1}^{N} \varepsilon_n. \tag{4}$$

This measure has been used in studies such as Färe and Grosskopf (1994, p. 103) and Orea (2002).

As demonstrated by Lovell (2003), Orea (2002), and Feng and Serletis (2010), the productivity index in (3) can be decomposed into three components: technical change ($TC$), change in technical efficiency ($\Delta TE$), and scale effects ($SC$):

$$\frac{d \ln TFP}{dt} = TC + \Delta TE + SC, \tag{5}$$

where

$$\begin{aligned} TC &= -\partial \ln D_o\left(\boldsymbol{y}, \boldsymbol{x}, t\right) / \partial t; \\ \Delta TE &= -\partial \ln \psi(t) / \partial t; \\ SC &= (\varepsilon - 1) \sum_{n=1}^{N} \left(-\frac{\varepsilon_n}{\varepsilon}\right) \dot{x}_n. \end{aligned} \tag{6}$$

The first term ($TC$) in (5) is a primal measure of the rate of technical change, capturing the shift in the best practice distance frontier. The second term ($\Delta TE$) is a primal measure of the change in technical efficiency, representing the rate at which an observed firm moves towards or away from the frontier. The third term ($SC$) captures the contribution of scale economies.

## 3. The Translog Stochastic Distance Frontier (SDF) Model with Time-Varying Heterogeneity

To estimate the productivity and efficiency measures in Section 2, it is necessary to parameterize the output distance function. In this paper we choose a translog functional form for the output distance function, mainly because it is easy to impose the linear homogeneity property with this functional form. Studies that have employed a translog output distance function include Färe *et al.* (1993) and O'Donnell and Coelli (2005). However, the standard translog output distance function has two drawbacks: first, it is inestimable because $D_o(\boldsymbol{y}, \boldsymbol{x}, t)$ is unobservable; and second, it does not allow for unobserved heterogeneity. To overcome these two drawbacks, we first follow Lovell *et al.* (1994) and O'Donnell and Coelli (2005) and transform the standard translog output distance function into an estimable regression equation in the form of a standard stochastic frontier model. We then follow Ahn *et al.* (2013) and use a multi-factor structure to model time-varying heterogeneity. These two steps will transform the standard translog output distance function into our translog SDF model with time-varying heterogeneity.

We start by specifying the standard translog output distance function as follows

$$\begin{aligned} \ln D_o(\boldsymbol{y}, \boldsymbol{x}, t) = a_0 &+ \sum_{m=1}^{M} a_m \ln y_m + \frac{1}{2} \sum_{m=1}^{M} \sum_{p=1}^{M} a_{mp} \ln y_m \ln y_p \\ &+ \sum_{n=1}^{N} b_n \ln x_n + \frac{1}{2} \sum_{n=1}^{N} \sum_{j=1}^{N} b_{nj} \ln x_n \ln x_j + \delta_\tau t + \frac{1}{2} \delta_{\tau\tau} t^2 \\ &+ \sum_{n=1}^{N} \sum_{m=1}^{M} g_{nm} \ln x_n \ln y_m + \sum_{m=1}^{M} \delta_m t \ln y_m + \sum_{n=1}^{N} \rho_n t \ln x_n, \end{aligned} \tag{7}$$

where $t$ denotes a time trend. The usual symmetry restrictions require $a_{mp} = a_{pm}$ and $b_{nj} = b_{jn}$. Moreover, to ensure linear homogeneity of the output distance function in $\boldsymbol{y}$, the following restrictions are imposed

$$\sum_{m=1}^{M} a_m = 1;\ \sum_{p=1}^{M} a_{mp} = 0;\ \sum_{m=1}^{M} g_{nm} = 0;\ \sum_{m=1}^{M} \delta_m = 0. \tag{8}$$

As mentioned above, a problem with (7) is that it is inestimable. This problem can be overcome by exploiting the linear homogeneity restrictions in (8). Specifically, we follow Lovell *et al.* (1994) and O'Donnell and Coelli (2005) and impose the linear homogeneity by normalizing (7) by one of the outputs (say, output $M$)

$$\begin{aligned}
\ln D_o\left(\frac{\boldsymbol{y}}{y_M}, \boldsymbol{x}, t\right) &= \ln\left[\frac{1}{y_M} D_o(\boldsymbol{y}, \boldsymbol{x}, t)\right] \\
&= -\ln y_M + \ln\left[D_o(\boldsymbol{y}, \boldsymbol{x}, t)\right] \\
&= -\ln y_M - \ln(\psi) \\
&= -\ln y_M - u,
\end{aligned} \tag{9}$$

where the first equality is obtained by the linear homogeneity property, the third one by (2). $u \equiv \ln(\psi) = -\ln D_o(\boldsymbol{y}, \boldsymbol{x}, t) \geq 0$ is a measure of inefficiency that is unobservable and non-negative. Rearranging (9) yields

$$-\ln y_M = \ln D_o\left(\frac{\boldsymbol{y}}{y_M}, \boldsymbol{x}, t\right) + u. \tag{10}$$

Assuming that $u$ follows a non-negative distribution and adding an independently and identically normally distributed error term, $v$, (10) can be further written as

$$-\ln y_M = \ln D_o\left(\frac{\boldsymbol{y}}{y_M}, \boldsymbol{x}, t\right) + u + v. \tag{11}$$

The above procedure thus transforms the standard translog output distance function in (7) into (11), an estimable equation in the form of a standard stochastic frontier model. However, like other standard stochastic frontier models, (11) does not allow for unobserved heterogeneity. In this sense, we refer to (11) as the standard translog stochastic distance frontier (SDF) model.

The standard translog SDF model in (11) can be written more explicitly by expanding the first term on the right hand side

$$
\begin{aligned}
-\ln y_M = a_0 &+ \sum_{m=1}^{M-1} a_m \ln\left(\frac{y_m}{y_M}\right) + \frac{1}{2}\sum_{m=1}^{M-1}\sum_{p=1}^{M-1} a_{mp} \ln\left(\frac{y_m}{y_M}\right)\ln\left(\frac{y_p}{y_M}\right) \\
&+ \sum_{n=1}^{N} b_p \ln x_p + \frac{1}{2}\sum_{n=1}^{N}\sum_{j=1}^{N} b_{nj} \ln x_n \ln x_j + \delta_\tau t + \frac{1}{2}\delta_{\tau\tau} t^2 \\
&+ \sum_{n=1}^{N}\sum_{m=1}^{M-1} g_{nm} \ln x_n \ln\left(\frac{y_m}{y_M}\right) + \sum_{m=1}^{M-1} \delta_m t \ln\left(\frac{y_m}{y_M}\right) + \sum_{n=1}^{N} \rho_n t \ln x_n + u + v. \qquad (12)
\end{aligned}
$$

In matrix notations, (12) can be written as

$$q_{it} = \boldsymbol{z}'_{it}\boldsymbol{\beta} + u_{it} + v_{it}, \qquad (13)$$

where $i = 1, \cdots, K$ indexes firms; $t = 1, \cdots, T$ indexes time; $q_{it} = -\ln y_{M,it}$; $\boldsymbol{z}_{it}$ is a vector comprising all the variables on the right hand side of (12); and $\boldsymbol{\beta}$ refers to the corresponding vector of coefficients of the translog function (including the intercept). In addition, $u_{it}$ ∼i.i.d. $\exp(\lambda^{-1})$ with '$\exp$' denoting an exponential distribution with an unknown parameter $\lambda$, and $v_{it}$ ∼i.i.d. $N(0, \sigma_v^2)$.

We now turn to introducing time-varying unobserved heterogeneity into (13). Specifically, we follow Ahn *et al*. (2013) and use a "multi-factor structure" to capture time-varying unobserved heterogeneity. Formally, the translog SDF model with time-varying heterogeneity can be written as follows

$$
\begin{aligned}
q_{it} &= \boldsymbol{z}'_{it}\boldsymbol{\beta} + \boldsymbol{f}'_t\boldsymbol{\gamma}_i + u_{it} + v_{it} \\
&= \boldsymbol{z}'_{it}\boldsymbol{\beta} + (f_{t1}\gamma_{i1} + f_{t2}\gamma_{i2} + ... + f_{th}\gamma_{ih}) + u_{it} + v_{it}, \qquad (14)
\end{aligned}
$$

where $u_{it}$ and $v_{it}$ are specified as above; $\boldsymbol{\gamma}_i = (\gamma_{i1}, ..., \gamma_{ih})'$ is an $h \times 1$ vector of unobserved firm-specific variables; and $\boldsymbol{f}_t = (f_{t1}, ..., f_{th})'$ is the corresponding vector of parameters measuring the effects of $\boldsymbol{\gamma}_i$ on $q_{it}$. For identification purpose, it is required that $h \leq (T-1)/2$ (Geweke and Zhou, 1996, p. 565)[2]. With regards to the specifications of $\boldsymbol{f}_t$ and $\boldsymbol{\gamma}_i$, we follow Geweke and Zhou (1996) and treat both as random. This treatment has two advantages. First, it facilitates the estimation of the model within a Bayesian framework, as can be seen below. Second, it does not impose any parametric form on $\boldsymbol{f}_t$, thus giving the model flexibility in capturing time variations of unobserved heterogeneity.

The translog SDF model with time-varying heterogeneity in (14) is very general. It reduces to the standard translog SDF model in (13) when $\boldsymbol{f}_t \equiv 0$ (i.e., in the absence of unobserved heterogeneity). It also reduces to the following true random effect SDF model when $h$ (the number of factors) is one and $\boldsymbol{f}_t$ is a constant (i.e., when the unobserved heterogeneity is time invariant)

$$q_{it} = \boldsymbol{z}'_{it}\boldsymbol{\beta} + w_i + u_{it} + v_{it}. \tag{16}$$

where $w_i \sim$i.i.d. $N(0, \sigma_w^2)$ represents time-invariant heterogeneity. For convenience, we refer to (16) as the translog SDF model with time-invariant heterogeneity.

It is straightforward to show that for the translog SDF model with time-varying heterogeneity,

[2] If (14) represents a panel data model with common factors where some factors are observable, (14) can be written as

$$\begin{aligned} q_{it} &= \boldsymbol{z}'_{it}\boldsymbol{\beta} + \boldsymbol{f}'_{1,t}\boldsymbol{\gamma}_{1,i} + \boldsymbol{f}'_{2,t}\boldsymbol{\gamma}_{2,i} + u_{it} + v_{it} \\ &= \widetilde{\boldsymbol{z}}'_{it}\widetilde{\boldsymbol{\beta}}_i + \boldsymbol{f}'_{2,t}\boldsymbol{\gamma}_{2,i} + u_{it} + v_{it}, \end{aligned} \tag{15}$$

where $\boldsymbol{f}_{1,t}$ is a $h_1 \times 1$ vector of unobservables; $\boldsymbol{f}_{2,t}$ is a $h_2 \times 1$ vector of unobservables; $\widetilde{\boldsymbol{z}}_{it} = \left(\boldsymbol{z}_{it}, \boldsymbol{f}_{1,t}\right)'$; and $\widetilde{\boldsymbol{\beta}}_i = \left(\boldsymbol{\beta}', \boldsymbol{\gamma}'_{1,i}\right)'$. (15) is a random coefficient model with a new factor structure, represented by $\boldsymbol{f}'_{2,t}\boldsymbol{\gamma}_{2,i}$. The identification restriction thus becomes $h_2 \leq (T-1)/2$. Accordingly, the prior and posterior distribution for $\widetilde{\boldsymbol{\beta}}_i$ needs to be changed. But, our specifications and discussions regarding the new factor structure remain the same. Further, if $f_{1,t}$ is a constant scalar (say $f_1$), the above model reduces to

$$q_{it} = \boldsymbol{z}'_{it}\boldsymbol{\beta} + w_i + \boldsymbol{f}'_{2,t}\boldsymbol{\gamma}_{2,i} + u_{it} + v_{it},$$

where $w_i = f_1\gamma_{1,i}$. The identification restriction is still $h_2 \leq (T-1)/2$.

technical efficiency, technical change, and returns to scale are respectively

$$
\begin{aligned}
TE &= \exp(-u); && (17)\\
TC &= -\frac{\partial \ln D_o\left(\boldsymbol{y},\boldsymbol{x},t\right)}{\partial t} \\
&= -\left(\delta_\tau + \delta_{\tau\tau}t + \sum_{m=1}^{M}\delta_m \ln y_m + \sum_{n=1}^{N}\rho_n \ln x_n\right); && (18)\\
RTS &= -\sum_{n=1}^{N}\frac{\partial \ln D_o\left(\boldsymbol{y},\boldsymbol{x},t\right)}{\partial \ln x_n}. && (19)
\end{aligned}
$$

Equation (17) can then be used to obtain efficiency change, $\Delta TE = -du/dt$.

## 4. Bayesian Estimation

In this section, we discuss a Bayesian procedure for estimating the translog SDF model with time-varying heterogeneity in (14). We first introduce some matrix notations: $\boldsymbol{q}_i = (q_{i1},...,q_{iT})'$, $\boldsymbol{q}= (\boldsymbol{q}'_1,...,\boldsymbol{q}'_K)'$, $\boldsymbol{q}_t = (q_{1t},...,q_{Kt})'$, $\boldsymbol{z}_i = (\boldsymbol{z}_{i1},...,\boldsymbol{z}_{iT})'$, $\boldsymbol{z}= (\boldsymbol{z}'_1,...,\boldsymbol{z}'_K)'$, $\boldsymbol{z}_t = (\boldsymbol{z}_{1t},...,\boldsymbol{z}_{Kt})'$, $\boldsymbol{u}_i = (u_{i1},...,u_{iT})'$, $\boldsymbol{u}= (\boldsymbol{u}'_1,...,\boldsymbol{u}'_K)'$, $\boldsymbol{u}_t = (u_{1t},...,u_{Kt})'$, $\boldsymbol{\Gamma}= (\boldsymbol{\gamma}_1,...,\boldsymbol{\gamma}_K)$, and $\boldsymbol{F}= (\boldsymbol{f}_1,...,\boldsymbol{f}_T)$.

Following Koop and Steel (2003) and O'Donnell and Coelli (2005), we adopt a flat prior for $\boldsymbol{\beta}$

$$p\left(\boldsymbol{\beta}\right) \propto 1, \tag{20}$$

and the following distribution for $h_v$

$$p(h_v) \propto \frac{1}{h_v}, \quad \text{where } h_v = \frac{1}{\sigma_v^2} > 0. \tag{21}$$

(21) implies that $h_v$ is fully determined by the likelihood function — see the conditional posterior distribution for $h_v$ in (29).

As mentioned above, we choose an exponential distribution for $u_{it}$ (O'Donnell and Coelli, 2005). Since the exponential distribution is a special case of the gamma distribution, the prior for

$u_{it}$ is written as

$$p\left(u_{it}\left|\lambda^{-1}\right.\right)=f_{\text{Gamma}}\left(u_{it}\left|1,\lambda^{-1}\right.\right), \tag{22}$$

where $f_{\text{Gamma}}$ denotes the gamma distribution.

According to Fernandez *et al.* (1997), in order to obtain a proper posterior we need a proper prior for the parameter, $\lambda$. We use the following proper prior

$$p(\lambda^{-1})=f_{\text{Gamma}}(\lambda^{-1}\left|1,-\ln\tau^{*}\right.), \tag{23}$$

where $\tau^{*}$ is the prior median of the efficiency distribution. Our best prior knowledge of the efficiency of BHCs in the U.S. is the mean efficiency value of 0.90 reported by Stiroh (2000) that examines the productivity and efficiency for BHCs in the U.S. over the period 1991–1997. We experimented with various values of $\tau^{*}$ ranging from $0.50$ to $0.99$. The results were always the same up to the number of digits presented in Section 6, suggesting that our results are very robust to large changes in $\tau^{*}$.

As stated above, we follow Geweke and Zhou (1996) and assume that $\boldsymbol{\gamma}_i$ and $\boldsymbol{f}_t$ follow the following distributions respectively

$$\boldsymbol{\gamma}_i\sim\text{i.i.d. } N(0,I_h), \text{ for } i=1,2,...,K; \tag{24}$$

and

$$p\left(\boldsymbol{f}_t\right)\propto 1, \text{ for } t=1,2,...,T. \tag{25}$$

The likelihood function can be shown to be

$$\begin{aligned} L\left(\boldsymbol{q}\left|\boldsymbol{\beta},h_v,\boldsymbol{u},\lambda^{-1},\boldsymbol{\Gamma},\boldsymbol{F}\right.\right) &= \prod_{i=1}^{K}\prod_{t=1}^{T}\left\{\sqrt{\frac{h_v}{2\pi}}\exp\left[-\frac{h_v}{2}\left(q_{it}-\boldsymbol{z}_{it}'\boldsymbol{\beta}-u_{it}-\boldsymbol{f}_t'\boldsymbol{\gamma}_i\right)^2\right]\right\} \\ &\propto h_v^{KT/2}\exp\left[-\frac{h_v}{2}\boldsymbol{v}'\boldsymbol{v}\right], \end{aligned} \tag{26}$$

where $\boldsymbol{v}=(v_{it})$, with $v_{it}=q_{it}-\boldsymbol{z}_{it}'\boldsymbol{\beta}-u_{it}-\boldsymbol{f}_t'\boldsymbol{\gamma}_i$.

Using Bayes's Theorem and combining the likelihood function in (26) and the prior distributions in (20) – (25), we obtain the following posterior joint density function

$$
\begin{aligned}
f\left(\boldsymbol{\beta}, h_v, \boldsymbol{u}, \lambda^{-1}, \boldsymbol{\Gamma}, \boldsymbol{F} \,|\, \boldsymbol{q}\right) \propto\; & h_v^{KT/2-1} \exp\left[-\frac{h_v}{2}\boldsymbol{v}'\boldsymbol{v}\right] \\
& \times \prod_{i=1}^{K}\prod_{t=1}^{T}\left[\lambda^{-1}\exp\left(-\lambda^{-1}u_{it}\right)\right]\exp\left(\lambda^{-1}\ln\tau^{*}\right) \\
& \times \prod_{i=1}^{K}\exp\left[-\frac{1}{2}\boldsymbol{\gamma}_i'\boldsymbol{\gamma}_i\right]. \qquad (27)
\end{aligned}
$$

Note that all the measures of productivity, technical change, technical efficiency, returns to scale are functions of $\boldsymbol{\beta}, h_v, \boldsymbol{u}, \lambda^{-1}, \boldsymbol{\Gamma}$, and $\boldsymbol{F}$. Let $g(\boldsymbol{\beta}, h_v, \boldsymbol{u}, \lambda^{-1}, \boldsymbol{\Gamma}, \boldsymbol{F})$ represent these functions of interest. In theory, we could obtain the moments of $g(\boldsymbol{\beta}, h_v, \boldsymbol{u}, \lambda^{-1}, \boldsymbol{\Gamma}, \boldsymbol{F})$ from the posterior density through integration. Unfortunately, these integrals cannot be computed analytically. Therefore, we use the Gibbs sampling algorithm which takes sequential random draws from a series of full conditional posterior distributions. Under very mild assumptions (Tierney, 1994), these draws converge to draws from the joint posterior. Once draws from the joint distribution are obtained, any posterior feature of interest can be calculated.

The full conditional posterior distributions for $\boldsymbol{\beta}, h_v, \boldsymbol{u}$, and $\lambda^{-1}$ are shown to be

$$
p\left(\boldsymbol{\beta} \,\middle|\, \boldsymbol{q}, h_v, \boldsymbol{u}, \lambda^{-1}, \boldsymbol{\Gamma}, \boldsymbol{F}\right) \propto f_{\text{Normal}}\left(\boldsymbol{\beta} \,\middle|\, \boldsymbol{b}, h_v^{-1}\left(\boldsymbol{z}'\boldsymbol{z}\right)^{-1}\right), \qquad (28)
$$

$$
p\left(h_v \,\middle|\, \boldsymbol{q}, \boldsymbol{\beta}, \boldsymbol{u}, \lambda^{-1}, \boldsymbol{\Gamma}, \boldsymbol{F}\right) \propto f_{\text{Gamma}}\left(h_v \,\middle|\, \frac{KT}{2}, \frac{1}{2}\boldsymbol{v}'\boldsymbol{v}\right), \qquad (29)
$$

$$
\begin{aligned}
p\left(\boldsymbol{u} \,\middle|\, \boldsymbol{q}, \boldsymbol{\beta}, h_v, \lambda^{-1}, \boldsymbol{\Gamma}, \boldsymbol{F}\right) \propto\; & f_{\text{Normal}}\left(\boldsymbol{u} \,|\, \boldsymbol{q} - \boldsymbol{z}\boldsymbol{\beta} - Fs - \left(h_v\lambda\right)^{-1}\boldsymbol{\iota}_{KT},\; h_v^{-1}\mathbf{I}_{KT}\right) \\
& \times \prod_{i=1}^{K}\prod_{t=1}^{T} I\left(\boldsymbol{u}_{it} \geq 0\right), \qquad (30)
\end{aligned}
$$

$$
p\left(\lambda^{-1} \,|\, \boldsymbol{q}, \boldsymbol{\beta}, h_v, \boldsymbol{u}, \boldsymbol{\Gamma}, \boldsymbol{F}\right) \propto f_{\text{Gamma}}\left(\lambda^{-1} \,|\, KT+1, \boldsymbol{u}'\boldsymbol{\iota}_{KT} - \ln\tau^{*}\right), \qquad (31)
$$

where $\boldsymbol{b} = (\boldsymbol{z}'\boldsymbol{z})^{-1}\boldsymbol{z}'[\boldsymbol{q}-\boldsymbol{u}-Fs]$; $Fs = \left(\boldsymbol{f}_1^{'}\gamma_1, ..., \boldsymbol{f}_T^{'}\gamma_1, ..., \boldsymbol{f}_1^{'}\gamma_K, ..., \boldsymbol{f}_T^{'}\gamma_K\right)'$; $\boldsymbol{\iota}_{KT}$ is a $KT \times 1$ vector of ones.

The conditional posterior densities of $\boldsymbol{\gamma}_i$, for $i = 1, 2, ..., K$, are multivariate normal as follows:

$$p\left(\boldsymbol{\gamma}_i \left| \boldsymbol{q}, \boldsymbol{\beta}, h_v, \boldsymbol{u}, \lambda^{-1}, \boldsymbol{F}\right.\right) \propto f_{\text{Normal}}\left(\overline{\gamma_i}, \boldsymbol{\Omega}_i\right), \tag{32}$$

where

$$\begin{aligned}
\overline{\gamma}_i &= \left(I_h + h_v \boldsymbol{F}\boldsymbol{F}'\right)^{-1} h_v \boldsymbol{F}\left(\boldsymbol{q}_i - \boldsymbol{z}_i\boldsymbol{\beta} - \boldsymbol{u}_i\right), \\
\boldsymbol{\Omega}_i &= \left(I_h + h_v \boldsymbol{F}\boldsymbol{F}'\right)^{-1}.
\end{aligned}$$

The full conditional posterior distributions for $\boldsymbol{F}$ is more involved than those given in (28) — (32). Specifically, for identification purpose (Geweke and Zhou, 1996), we set the first $h$ columns of $F = \left(\underset{h\times h}{\boldsymbol{F}_1}, \underset{h\times(T-h)}{\boldsymbol{F}_2}\right)$ to

$$\boldsymbol{F}_1 = \begin{pmatrix} f_{11} & f_{21} & ... & f_{h1} \\ 0 & f_{22} & ... & f_{h2} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & ... & f_{hh} \end{pmatrix}$$

with $f_{tt} > 0$ for $t = 1, ..., h$. Then the conditional posterior densities for $\boldsymbol{F}_1$ are as follows

$$p\left(\boldsymbol{f}_t \left| \boldsymbol{q}, \boldsymbol{\beta}, h_v, \boldsymbol{u}, \lambda^{-1}, \boldsymbol{\Gamma}\right.\right) \propto f_{\text{Normal}}\left(\overline{\boldsymbol{f}}_t, \boldsymbol{\Sigma}_t\right), \tag{33}$$

where

$$\begin{aligned}
\overline{\boldsymbol{f}}_t &= \left(\boldsymbol{\Gamma}_t\boldsymbol{\Gamma}_t^{'}\right)^{-1} \boldsymbol{\Gamma}_t\left(\boldsymbol{q}_t - \boldsymbol{z}_t\boldsymbol{\beta} - \boldsymbol{u}_t\right), \\
\boldsymbol{\Sigma}_t &= h_v^{-1}\left(\boldsymbol{\Gamma}_t\boldsymbol{\Gamma}_t^{'}\right)^{-1}, \\
\boldsymbol{\Gamma}_t &= \left(\boldsymbol{\gamma}_1^{(t)}, ..., \boldsymbol{\gamma}_K^{(t)}\right), \\
\boldsymbol{\gamma}_i^{(t)} &= \left(\gamma_{i1}, ..., \gamma_{it}\right)' (i = 1, ..., K).
\end{aligned}$$

And the conditional posterior densities for $\boldsymbol{F}_2$, i.e. $t = h+1, \ldots, T$, are given by

$$p\left(\boldsymbol{f}_t \,\middle|\, \boldsymbol{q}, \boldsymbol{\beta}, h_v, \boldsymbol{u}, \lambda^{-1}, \boldsymbol{\Gamma}\right) \propto N\left(\overline{\boldsymbol{f}}_t, \boldsymbol{\Sigma}_t\right), \tag{34}$$

where

$$\begin{aligned} \overline{\boldsymbol{f}}_t &= \left(\boldsymbol{\Gamma}\boldsymbol{\Gamma}'\right)^{-1} \boldsymbol{\Gamma}\left(\boldsymbol{q}_t - \boldsymbol{z}_t\boldsymbol{\beta} - \boldsymbol{u}_t\right), \\ \boldsymbol{\Sigma}_t &= h_v^{-1}\left(\boldsymbol{\Gamma}\boldsymbol{\Gamma}'\right)^{-1}. \end{aligned}$$

The Gibbs sampler for Bayesian estimation can then be implemented by drawing sequentially from the conditional posteriors in (28)–(34).

## 5. The Data

The data used in this study are obtained from the consolidated FR Y-9C reports filed by U.S. BHCs over the period 2004 – 2013. Studies (Cole *et al.*, 2004; Berger *et al.*, 2005) suggest that large and small BHCs/banks employ different production technologies. Specifically, large BHCs/banks tend to employ "hard" information-based production technologies, while small ones tend to employ "soft" information-based production technologies (Berger *et al.*, 2005). To avoid potential problems associated with technology heterogeneity, in this paper we focus on a selected subsample of relatively homogeneous large BHCs, namely those with total assets in excess of 1 billion dollars[3] (in 2004 U.S. dollars) in each of the first three years of the sample period. Due to exit by BHCs that were previously part of the sample, we end up with an unbalanced sample of 335 BHCs. The use of an unbalanced panel may mitigate potential survivorship bias caused by the omission of BHCs that did not survive[4].

To select the relevant variables, we follow the commonly-accepted intermediation approach proposed by Sealey and Lindley (1977), whereby banks collect purchased funds and use labor

[3] \$1 billion is widely accepted as a cutoff for separating large and small BHCs/banks (see, for example, Cole *et al.*, 2004).

[4] The use of a balanced panel might result in survivorship bias. However, we also note that the use of an unbalanced panel may potentially distort inter-temporal comparisons of banking sector efficiency.

and capital to transform these funds into loans and other assets. On the input side, three inputs are included: the quantity of labor, $x_1$; the quantity of purchased funds and deposits, $x_2$; and the quantity of physical capital, $x_3$, which includes premises and other fixed assets. On the output side, four outputs are specified. They are consumer loans, $y_1$; securities, $y_2$, which includes all non-loan financial assets (i.e., all financial assets minus the sum of all loans, securities, and equity); non-consumer loans, $y_3$, which is composed of industrial, commercial, and real estate loans; and off-balance sheet items, $y_4$. All the quantities are constructed by following Berger and Mester (2003), with the exception of $y_4$, which is constructed by following the User's Guide for the BHC Performance Report published by the Federal Reserve Board (2013). These quantities are deflated by the GDP deflator to the base year 2004, except for the quantity of labor.

## 6. Empirical Results

### *6.1. Model comparison*

In this subsection, we compare the estimation performance of the three SDF models, namely, the SDF model with time-varying heterogeneity in (14), the standard translog SDF model in (13) and the translog SDF model with time-invariant heterogeneity in (16).

Before making this comparison, we first discuss two things: i) the choice of $h$ (i.e., the number of factors) for the SDF model with time-varying heterogeneity, and ii) the convergence performance of the three SDF models. With regard to the choice of $h$, given that we have $T = 10$ time periods, the highest number of factors identifiable is $4$ (i.e. $h \leq (T-1)/2$). We, therefore, estimate four different versions of the SDF model with time-varying heterogeneity, with the first version having one factor, the second version having two factors, the third version having three factors, and the last version having four factors. Following Geweke *et al*. (2015), we use the marginal log likelihood to choose among the four versions. As can be seen from Table 1, the marginal log likelihood increases until $h = 4$, indicating that the SDF model with four factors is preferred. To check the robustness of this finding, we also calculate the deviance information criterion (Spiegelhalter *et al*., 2002) for each of the four versions. The idea behind DIC is that models with smaller

DIC should be preferred to models with larger DIC. The second row of Table 1 shows that the SDF model with four factors has the lowest DIC value, confirming that the model with four factors is preferred. Hence, in what follows we will use the SDF model with four factors wherever the SDF model with time-varying heterogeneity is needed.

To evaluate the convergence performance of the three SDF models, we calculate the simulation inefficiency factors (SIF) for each model. SIF can be interpreted as the number of successive iterations needed to obtain near independent draws (see, for example Kim *et al.*, 1998). In our experience, a sampler can achieve reasonable mixing performance when the resulting SIF value is below 100. As can be seen from Table 2.3, all SIF values for the standard translog SDF model are less than $20$, suggesting strongly that the sampler has converged. The SIF values for the translog SDF model with time-invariant heterogeneity (displayed in Table 2.2) are all less than $61$, indicating that the sampler for this model has also converged. The SIF values for the translog SDF model with time-varying heterogeneity (shown in Table 2.1) are all less than $47$, suggesting that the sampler for this latter model has converged too.

We now turn to comparing the three SDF models using the Bayes factor (Kass and Raftery 1995). Letting $M_I$ and $M_J$ denote two competing models, the Bayes factor is defined as the ratio of the posterior odds of $M_J$ to $M_I$ multiplied by the prior odds of $M_J$ to $M_I$. When both models have an equal prior likelihood, the Bayes factor reduces to

$$B_{JI} = \frac{\Pr\left(D\,|M_J\right)}{\Pr\left(D\,|M_I\right)},$$

where $\Pr\left(D\,|M_J\right)$ and $\Pr\left(D\,|M_I\right)$ represent the marginal likelihood (integrating over the model parameters) of the data $D$ for $M_J$ to $M_I$, respectively. The Bayes factor summarizes "the evidence provided by the data in favor of one scientific theory, represented by a statistical model, as opposed to another" (Kass and Raftery, 1995). Kass and Raftery suggest using the following Schwarz criterion to approximate the Bayes factor: $S = l\left(D\,|M_J\right) - l\left(D\,|M_I\right) - \frac{1}{2}\left(d_J - d_I\right)\log n \approx \ln B_{JI}$, where $l\left(\cdot\right)$ is the maximized log likelihood, $d$ is the number of model parameters, and $n$ is the sample size. $2 \times S$ can then be used with the following table to judge which model is preferred by the data

| $2 \ln B_{JI}$ | Evidence against $M_I$ |
|---|---|
| 0 to 2 | Not worth more than a bare mention |
| 2 to 6 | Positive |
| 6 to 10 | Strong |
| $>$10 | Very strong |

In this study, the value of $2S$ of the translog SDF model with time-varying heterogeneity against the standard translog SDF model is $900.78$, while that of the translog SDF model with time-invariant heterogeneity against the standard translog SDF model is $581.38$. This suggests strongly that the two models allowing for unobserved heterogeneity outperform the standard SDF model. In other words, unobserved heterogeneity is present among the BHCs. More importantly, the value of $2S$ of the translog SDF model with time-varying heterogeneity against the translog SDF model with time-invariant heterogeneity is $322.40$, providing strong evidence that the former model is favored over the latter. Put differently, unobserved heterogeneity varies over time in the U.S. banking industry.

To check the robustness of this finding, we also calculate and report DIC values and their 95% credible intervals for the three SDF models in Tables 2.1, 2.2, and 2.3, respectively. Both the DIC value for the translog SDF model with time-invariant heterogeneity ($-20715.33$) and that for the translog SDF model with time-varying heterogeneity ($-32824.69$) are significantly smaller than that for the standard translog SDF model ($-17533.76$). This confirms that unobserved heterogeneity is present among the BHCs. In addition, the DIC value for the translog SDF model with time-varying heterogeneity is much smaller than that for the translog SDF model with time-invariant heterogeneity, confirming that unobserved heterogeneity is time-varying in the U.S. banking industry.

Possible candidates for the four time-varying unobserved factors include asset quality, geographic reach and local branch density, organizational structure, and "too-big-to-fail" factors.

These four candidates have been widely documented to have important impacts on the production process of banks (see Hughes and Mester, 1993; Dick, 2006; Berger *et al*., 2005; Davies and Tracey, 2014). It should be noted that while these four factors are unknown to the econometrician, they are likely to be known to firm (bank) managers (see Reiss and Wolak, 2007). The time-varying nature of these four factors implies that the effects of these factors on the production process of banks vary over time. For instance, factors, which have helped increase revenues or reduce costs today, may be less helpful or even not helpful in the future. Therefore, bank managers should closely monitor the trends of these factors so that they have the current information on these factors when making production decisions. They should also take preventive and proactive measures so that these factors evolve in a way that leads to higher revenues or lower costs in the future.

*6.2. Consequences of failure to take into account unobserved heterogeneity*

It is of interest to briefly examine consequences of failing to allow for time-varying unobserved technology heterogeneity. Due to space limitations, we use productivity growth as an example, because productivity growth is a more comprehensive measure of productivity and efficiency than technical change, technical efficiency, and returns to scale.

We undertake this examination in two steps. In the first step, we examine consequences of completely ignoring unobserved heterogeneity on the magnitudes of productivity growth, by comparing the standard translog SDF model and the translog SDF model with time-varying heterogeneity. Panel A, Table 3 presents the mean differences in productivity growth between these two models. As can be seen, the mean differences in productivity growth range between $-0.41\%$ and $0.56\%$ with an average of $0.14\%$. This means that by completely ignoring unobserved technology heterogeneity, the standard translog SDF model on average overestimates productivity growth by $0.14\%$. This latter figure is not small, considering that the average productivity growth over the sample period based on the translog SDF model with time-varying heterogeneity is only one-seventh of this figure (i.e., $0.02\%$; see Table 7).

We now turn to the second step. In this step we examine consequences of not allowing technology heterogeneity to vary over time, by comparing the translog SDF model with time-invariant heterogeneity and the translog SDF model with time-varying heterogeneity. Panel B, Table 3 presents the mean differences in productivity growth between these two models. As can be seen, the mean differences in productivity growth between these two models range between $-0.51\%$ and $-0.08\%$ with an average of $-0.35\%$. This suggests that by failing to allow technology heterogeneity to vary over time, the translog SDF model with time-invariant heterogeneity underestimates productivity growth by $-0.35\%$. Again, this figure is not small, because the average productivity growth over the sample period based on the translog SDF model with time-varying heterogeneity is only $0.02\%$ (see Table 7).

Hence, in what follows we concentrate on the empirical results from the translog SDF model with time-varying heterogeneity.

### *6.3. Empirical results from the SDF model with time-varying heterogeneity*

#### *6.3.1. Technical change*

Table 4 presents point estimates of average annual technical change and their 95% credible intervals. As can be seen, the BHCs have experienced small but positive rates of technical change, with an average of $0.16\%$. Our estimates of technical change are generally smaller than those found in Feng and Serletis (2010) and Feng and Zhang (2012). There are two possible reasons for this. First, the two previous studies only include commercial banks that operated continuously over the sample period, whereas our unbalanced panel not only includes BHCs that operated continuously over the sample period, but also less efficient BHCs that did not survive. Second, our measure of technical change, by definition, only includes pure technical change (i.e., technical change captured by the time trend). Technical change due to factors, such as improvements in output/input quality and branch networks, is captured by the multi-factor structure and thus excluded from our estimates. As pointed out by Hulten (1992), technical change is “a measure of our ignorance” and that the more we can explain it, the smaller pure technical change becomes.

#### *6.3.2. Technical efficiency*

Table 5.1 reports point estimates of average annual technical efficiency and their 95% credible intervals. Two findings stand out from this table. First, the BHCs as a whole operate at high levels of efficiency ranging from $96.86\%$ to $97.98\%$. Second, the average technical efficiency level has remained relatively stable over time, fluctuating within a narrow range between $96.86\%$ and $97.98\%$.

To get a better understanding of the distribution of technical efficiency across BHCs, Table 5.2 reports the distribution (including median, minimum, maximum, lower quartile, upper quartile) of the estimates of technical efficiency across BHCs for each year. As can be seen, technical efficiency differs substantially among the BHCs. Taking 2013 as an example, the highest technical efficiency is $99.79\%$, whereas the lowest is only $64.72\%$. In addition, the table shows that while the highest technical efficiency level has remained relatively stable over time, the lowest technical efficiency level has declined after the recent financial crisis, falling from $79.48\%$ in 2008 to $60.52\%$ in $2012$. This result suggests that the BHCs were not affected by the recent financial crisis to the same extent; instead, some were more affected than others.

*6.3.3 Returns to scale*

Table 6.1 presents point estimates of returns to scale (RTS) and their 95% credible intervals. It shows that for the subperiod 2004 – 2009 both the point estimates and credible intervals are slightly above one, indicating that on average the BHCs exhibit slight increasing returns to scale. However, for the subperiod 2010 – 2013 while the point estimates are slightly above one, their credible intervals all contain one. This indicates that we cannot reject that on average the BHCs show constant returns to scale for the second subperiod.

As with the case of technical efficiency, we are also interested in estimates of RTS at individual BHC level. We compute the percentage of BHCs facing increasing, constant, or decreasing returns to scale for each year. As in Wheelock and Wilson (2012), this computation is performed by counting the number of cases where the 95% credible intervals are strictly greater than 1.0 (indicating increasing returns to scale), strictly less than 1.0 (indicating decreasing returns to scale), or contain 1.0 (indicating constant returns to scale). The results are presented in Table 6.2. Two findings emerge from this table. First, on average more than half ($53.78\%$) of the BHCs face increasing returns to scale, $29.59\%$ of the BHCs face constant returns to scale, and $16.63\%$ of the BHCs face decreasing returns to scale. Second, the percentage of BHCs facing increasing returns to scale decreases markedly from $67.46\%$ in 2005 to $38.99\%$ in 2013. In contrast, the percentage of BHCs facing constant returns to scale increases significantly from $19.10\%$ in 2005 to $43.12\%$ in 2013, and that facing decreasing returns to scale fluctuates between $11.62\%$ – $21.40\%$.

It is of interest to examine if the differences in returns to scale across banks can be attributable to size change (or if there is any pattern between returns to scale and bank size). For this purpose, we regress the estimated returns to scale on a constant and total assets for each year, i.e. $RTS = \overline{\beta}_0 + \overline{\beta}_1 ASSET$, where $ASSET$ denotes total assets, $\overline{\beta}_0$ a constant, and $\overline{\beta}_1$ the coefficient for *ASSET*. For each year, we cannot reject that $\overline{\beta}_1 = 0$ at the 5% level of significance, indicating that there is no linear relationship between returns to scale and bank size. To further examine if there is a nonlinear relationship between asset size and returns to scale, we regress the estimated returns to scale on a constant, total assets, and squared total assets for each year, i.e. $RTS = \overline{\theta}_0 + \overline{\theta}_1 ASSET + \overline{\theta}_2 ASSET^2$, where $\overline{\theta}_0$ is a constant, $\overline{\theta}_1$ the coefficient for *ASSET*, and $\overline{\theta}_2$ the coefficient for $ASSET^2$ . Again for each year, we cannot reject that $\overline{\theta}_1 = \overline{\theta}_2 = 0$ at the 5% level of significance, confirming that there is no clear pattern between returns to scale and asset size. This finding is consistent with those of Daniel *et al.* (1973), Murray and White (1983), and Feng and Zhang (2014) Specifically, those studies find that when technology heterogeneity is accounted for, the commonly found pattern (i.e., banks face increasing returns to scale up to an optimum size and then decreasing returns to scale above that point) may not hold. In addition, we also run similar regressions for technical efficiency. Again, we find that there is no clear pattern between technical efficiency and asset size.

We also note that there are some similarities and some differences between the results presented here and those of other recent studies (Wheelock and Wilson, 2012; Hughes and Mester, 2013; Davies and Tracey, 2014). For example, while both this study and Wheelock and Wilson (2012) find evidence of increasing returns to scale for most banks for the period 2004 – 2006 (i.e., the period that overlaps in these two studies), the latter study finds a much higher percentage ($99.7\%$)

of BHCs facing increasing returns to scale, as compared with $60.69\%$ in this study. A possible reason for the differences is that different sample compositions and time periods are used. For example, this study uses an unbalanced panel of large U.S. BHCs with at least $1 billion of total assets over the period 2004 – 2013; whereas Wheelock and Wilson (2012) use an unbalanced panel of all top-tier BHCs and commercial banks in the U.S. over the period 1984 – 2006.

*6.3.4 TFP growth and its components*

Table 7 presents the results on productivity growth and its three components (i.e., technical change, efficiency change, and scale effects). It should be noted that the first year of the sample period is dropped because we need to difference technical efficiencies in two consecutive years to obtain efficiency changes. This table shows that the estimates of productivity growth are quite small, ranging between $-0.32\%$ and $0.49\%$.

We now turn to the three components of productivity growth. The estimates of scale effects are very small, ranging between $-0.02\%$ and $0.11\%$ (see the last panel of Table 7). This small magnitude is not surprising, given that the BHCs on average show slight increasing or constant returns to scale through the sample period. In addition, the estimates of scale effects are positive for all sample years with the exception of 2009 and 2010. This is because input reductions occurred in these two years (see equation (6)). As for the estimates of technical change presented in the third panel, they are small but positive for all sample years, as discussed previously. With regard to the estimates of efficiency change reported in the second panel, they are small and negative for all years except 2006 and 2009.

As with the case of technical change, there are two possible reasons for the small magnitude of productivity growth. First, our unbalanced panel not only includes surviving BHCs, but also less productive BHCs that did not survive. Second, our estimates of productivity growth only include pure technical change. Technical change due to factors, such as improvements in output/input quality and branch networks, is captured by the multi-factor structure and thus excluded from our estimates of productivity growth.

## 7. Conclusion

Productivity and efficiency of the U.S. banking industry has attracted much attention in the past three decades. One important line of research in this area focuses on how to deal with the problem of unobserved heterogeneity among banks. The present study aims to contribute to this line of research by applying a stochastic distance frontier model that allows for unobserved time-varying heterogeneity to bank holding companies (BHCs) in the U.S. This study is interesting because there is considerable evidence that time-varying unobserved heterogeneity is widely present in the industry.

Specifically, we estimate a translog stochastic distance frontier (SDF) model with time-varying heterogeneity. This model is obtained by adding time-varying individual effects to the standard translog SDF model. Following Ahn *et al*. (2013), these time-varying individual effects are modeled by a “multi-factor structure”. This model is very general. It reduces to the standard translog SDF model when there is no unobserved heterogeneity and to the ‘true random effect SDF model’ when there is only one time-invariant heterogeneity component.

Our empirical results show that the translog SDF model with time-varying heterogeneity outperforms both the standard SDF model and the true random effect SDF model, providing evidence for the presence of time-varying heterogeneity among BHCs in the U.S. Our analysis further demonstrates that failure to allow for time-varying unobserved heterogeneity results in misleading results regarding productivity and efficiency. Our results from the translog SDF model with time-varying heterogeneity show that the majority of large BHCs in the U.S. exhibit increasing returns to scale, a small percentage exhibit constant returns to scale, and an even smaller percentage exhibit decreasing returns to scale. They also show that on average the BHCs have experienced small positive or even negative technical change and productivity growth.

## References

Ahn, S.C., Lee, Y.H., Schmidt, P., 2007. Stochastic frontier models with multiple time-varying individual effects. Journal of Productivity Analysis 27, 1-12.

Ahn, S.C., Lee, Y.H., Schmidt, P., 2013. Panel data models with multiple time-varying individual effects. Journal of Econometrics 174(1), 1-14.

Berger, A.N., Mester, L.J., 2003. Explaining the dramatic changes in the performance of U.S. banks: Technological change, deregulation, and dynamic changes in competition. Journal of Financial Intermediation 12, 57–95.

Berger, A.N., Miller, N.H., Petersen, M.A., Rajan, R.G., Stein, J.C., 2005. Does function follow organizational form? Evidence from the lending practices of large and small banks. Journal Financial Economics 76, 237–269.

Caves, D.W., Christensen, L.R., Diewert, W.E.,1982. The economic theory of index numbers and the measurement of input, output, and productivity. Econometrica 50, 1393-1414.

Cole, R.A., Goldberg, L.G., White, L.J., 2004. Cookie cutter vs. character: the micro structure of small business lending by large and small banks. Journal Financial and Quantitative Analysis 39, 227–251.

Daniel, D.L., Longbrake, W.A., Murphy, N.B., 1973. The effect of technology on bank economies of scale for demand deposits. Journal of Finance 28, 31–146.

Demsetz, R .S., Strahan, P. E., 1997. Diversification, size, and risk at bank holding companies. Journal of Money, Credit and Banking 29(3), 300-313.

Dick, A. A. 2006. Nationwide branching and its impact on market structure, quality, and bank performance. The Journal of Business 79(2), 567-592.

Davies, R., Tracey, B., 2014. Too big to be efficient? The impact of implicit subsidies on estimates of scale economies for banks. Journal of Money, Credit and Banking 46 (1), 219-253.

El-Gamal, M.A., Inanoglu, H., 2005. Inefficiency and heterogeneity in Turkish banking: 1990–2000. Journal of Applied Economics 5, 641–664.

Färe, R., Grosskopf, S., 1994. Cost and Revenue Constrained Production. Springer-Verlag New York, Inc.

Färe, R., Primont, D., 1995. Multi-Output Production and Duality: Theory and Applications. Netherlands: Kluwer Academic Publishers.

Färe, R. Grosskopf, S., Lovell, C. A. K., 1993. Derivation of Shadow Prices for Undesirable Outputs: A Distance Function Approach. Review of Economics and Statistics 75(2), 374-380.

Federal Reserve Board Annual Report, 2012. http://www.federalreserve.gov/publications/annual-report/2012-contents.htm.

Federal Reserve Board, 2013. A User's Guide for the Bank Holding Company Performance Report, http://www.federalreserve.gov/boarddocs/supmanual/bhcpr/UsersGuide13/default.htm.

Feng, G., Serletis, A., 2010. A primal Divisia technical change index based on the output distance function. Journal of Econometrics 159, 320-330.

Feng, G., Zhang, X., 2012. Productivity and efficiency at large and community banks in the U.S.: a Bayesian true random effects stochastic distance frontier analysis. Journal of Banking and Finance 36 (7), 1883–1895.

Feng, G., Zhang, X., 2014. Returns to scale at large banks in the US: A random coefficient stochastic frontier approach. Journal of Banking and Finance 39, 135–145.

Fernandez, C., Osiewalski, J., Steel, M.F.J., 1997. On the use of panel data in stochastic frontier models with improper priors. Journal of Econometrics 79, 169-193.

Fuss, M., 1994. Productivity growth in Canadian telecommunications. Canadian Journal of Economics 27(2), 371-392.

Geweke, J., Zhou, G., 1996. Measuring the pricing error of the arbitrage pricing theory. Review of

Financial Studies 9(2), 557-587.

Geweke, J., Durham, G., Xu, H., 2015. Bayesian inference for logistic regression models using sequential posterior simulation, in: S.K. Upadhyay, U. Singh, D. K. Dey and A. Loganathan (eds.), Current Trends in Bayesian Methodology with Applications, Chapter 14, pp. 289 - 312.

Hannan, T.H., Hanweck, G.A., 1998. Bank insolvency risk and the market for large certificates of deposit. Journal of Money Credit and Banking 20(2), 203-211.

Hirtle, B., 2007. The impact of network size on bank branch performance. Journal of Banking and Finance 31, 3782–3805.

Hughes, J. P., Mester, L.J., 1993. A quality and risk-adjusted cost function for banks: Evidence on the "too-big-to-fail" doctrine. Journal of Productivity Analysis 4, 293-315.

Hughes, J. P., Mester, L.J., 1998. Bank capitalization and cost: Evidence of scale economies in risk management and signaling. Review of Economics and Statistics 80(2), 314-325.

Hughes, J. P., Mester, L.J., 2013. Who said large banks don't experience scale economies? Evidence from a risk-return driven cost function. Journal of Financial Intermediation 22, 559-585.

Hulten, C. R., 1992. Growth accounting when technical change is embodied in capital. The American Economic Review 82(4), 964-980.

Kass, R., Raftery, A., 1995, Bayes factors and model uncertainty. Journal of the American Statistical Association 90, 773–795.

Kim, S., Shephard, N., Chib, S., 1998. Stochastic volatility: likelihood inference and comparison with ARCH models. Review of Economic Studies 65, 361–393.

Koop, G., Steel, M., 2003. Bayesian analysis of stochastic frontier models. In: A Companion to Theoretical Econometrics, Baltagi, B. (eds.), MA: Blackwell.

Lovell, C.A.K., 2003, The decomposition of Malmquist productivity indexes. Journal of Productivity Analysis 20(3), 437 - 458.

Lovell, C.A.K., Richardson, S., Travers, P., Wood, L.L., 1994. Resources and functionings: A new view of inequality in Australia. In: Models and Measurement of Welfare and Inequality, W. Eichhorn (eds.), Berlin: Springer-Verlag Press, 787-807.

Mester, L.J., 1997. Measuring efficiency at US banks: accounting for heterogeneity is important. European Journal of Operational Research 98, 230–242.

Murray, J.D., White, R.W., 1983. Economies of scale and economies of scope in multiproduct financial institutions: a study of British Columbia credit unions. Journal of Finance 38, 887–902.

O'Donnell, C.J., Coelli, T.J., 2005. A Bayesian approach to imposing curvature on distance functions. Journal of Econometrics 126, 493-523.

Orea, L., 2002. Parametric decomposition of a generalized malmquist productivity index. Journal of Productivity Analysis 18, 5-22.

Reiss, P.C., Wolak, F.A., 2007. Structural econometric modeling: Rationales and examples from industrial organization. In: Handbook of Econometrics, Heckman, J.J., Leamer, E.E. (eds), North-Holland, 4277-4412.

Rosen, R.J., 2003. Is three a crowd? Competition among regulators in banking. Journal of Money, Credit and Banking 35, 967–998.

Rossi, C.V., 1998. Mortgage banking cost structure: resolving an enigma. Journal of Economics and Business 50, 219–234.

Sealey, C., Lindley, J., 1977. Inputs, outputs, and a theory of production and cost at depository financial institutions. Journal of Finance 32, 1251-1266.

Spiegelhalter, D.J., Best, N.G., Carlin B.P., van der Linde, A., 2002. Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society, Series B, 64, 583–640.

Stiroh, K.J., 2000. How did bank holding companies prosper in the 1990s? Journal of Banking and Finance 24, 1703-1745.

Tierney, L., 1994. Markov chains for exploring posterior distributions (with discussion). Annals of Statistics 22, 1701-1762.

Wheelock, D.C., Wilson, P.W., 2012. Do large banks have lower costs? New estimates of returns to scale for U.S. banks. Journal of Money, Credit and Banking 44 (1), 171-199.

Table 1

Selection on Number of Factors

| | 1 factor | 2 factors | 3 factors | 4 factors |
|---|---|---|---|---|
| log marginal likelihood | 10047.6 | 10163.2 | 10260.0 | 10340.7 |
| DIC | −30343.4 | −32187.8 | −32712.7 | −32824.7 |

Table 2.1

Parameter Estimates for The Translog SDF Model with Time-variant Heterogeneity

| Parameter | Estimate | 95% Credible Interval | SIF |
|---|---|---|---|
| $a_1$ | 0.0440 | $(0.0412, 0.0471)$ | 32.4379 |
| $a_2$ | 0.5339 | $(0.5263, 0.5412)$ | 35.4716 |
| $a_3$ | 0.4184 | $(0.4132, 0.4236)$ | 37.8097 |
| $a_4$ | 0.0037 | $(-0.0023, 0.0093)$ | 32.1027 |
| $a_{11}$ | 0.0069 | $(0.0061, 0.0078)$ | 46.0176 |
| $a_{12}$ | 0.0044 | $(0.0019, 0.0071)$ | 36.7598 |
| $a_{13}$ | $-0.0123$ | $(-0.0144, -0.0101)$ | 37.6129 |
| $a_{14}$ | 0.0009 | $(-0.0011, 0.0029)$ | 39.5101 |
| $a_{22}$ | 0.1768 | $(0.1696, 0.1843)$ | 20.6601 |
| $a_{23}$ | $-0.1769$ | $(-0.1821, -0.1721)$ | 31.3136 |
| $a_{24}$ | $-0.0043$ | $(-0.0092, 0.0008)$ | 26.3584 |
| $a_{33}$ | 0.1714 | $(0.1675, 0.1758)$ | 36.9931 |
| $a_{34}$ | 0.0178 | $(0.0138, 0.0217)$ | 28.2545 |
| $a_{44}$ | $-0.0144$ | $(-0.0205, -0.0088)$ | 36.8776 |
| $b_1$ | $-0.0011$ | $(-0.0148, 0.0114)$ | 40.6111 |
| $b_2$ | $-0.9853$ | $(-0.9965, -0.9733)$ | 39.2007 |
| $b_3$ | $-0.0149$ | $(-0.0229, -0.0070)$ | 35.5931 |
| $b_{11}$ | $-0.0943$ | $(-0.1175, -0.0729)$ | 41.9517 |
| $b_{12}$ | 0.1184 | $(0.1016, 0.1354)$ | 36.2693 |
| $b_{13}$ | $-0.0101$ | $(-0.0209, 0.0008)$ | 32.2884 |
| $b_{22}$ | $-0.1030$ | $(-0.1182, -0.0877)$ | 30.7076 |
| $b_{23}$ | $-0.0236$ | $(-0.0320, -0.0150)$ | 36.9559 |
| $b_{33}$ | 0.0235 | $(0.0156, 0.0310)$ | 24.0570 |
| $g_{11}$ | 0.0222 | $(0.0183, 0.0261)$ | 39.7789 |
| $g_{12}$ | $-0.0840$ | $(-0.0955, -0.0730)$ | 22.4906 |
| $g_{13}$ | 0.0437 | $(0.0356, 0.0519)$ | 31.7768 |
| $g_{14}$ | 0.0180 | $(0.0090, 0.0275)$ | 32.0237 |
| $g_{21}$ | $-0.0218$ | $(-0.0248, -0.0185)$ | 41.7147 |
| $g_{22}$ | 0.0841 | $(0.0751, 0.0932)$ | 20.2244 |
| $g_{23}$ | $-0.0544$ | $(-0.0598, -0.0487)$ | 27.2558 |
| $g_{24}$ | $-0.0079$ | $(-0.0156, -0.0008)$ | 30.0257 |
| $g_{31}$ | $-0.0046$ | $(-0.0070, -0.0022)$ | 27.6269 |
| $g_{32}$ | 0.0094 | $(0.0048, 0.0143)$ | 24.8913 |
| $g_{33}$ | $-0.0050$ | $(-0.0094, -0.0011)$ | 34.8711 |
| $g_{34}$ | 0.0002 | $(-0.0043, 0.0042)$ | 24.1695 |
| $\delta_1$ | $-0.0003$ | $(-0.0008, 0.0001)$ | 43.1827 |
| $\delta_2$ | 0.0008 | $(-0.0003, 0.0019)$ | 27.7825 |
| $\delta_3$ | $-0.0016$ | $(-0.0024, -0.0009)$ | 32.3954 |
| $\delta_4$ | 0.0011 | $(0.0002, 0.0020)$ | 24.7878 |
| $\delta_\tau$ | $-0.0001$ | $(-0.0007, 0.0004)$ | 17.8440 |
| $\delta_{\tau\tau}$ | 0.0000 | $(-0.0003, 0.0003)$ | 22.0423 |
| $\rho_1$ | $-0.0036$ | $(-0.0058, -0.0015)$ | 34.8079 |
| $\rho_2$ | 0.0010 | $(-0.0006, 0.0027)$ | 33.1017 |
| $\rho_3$ | 0.0028 | $(0.0017, 0.0039)$ | 31.1992 |
| DIC | $-32824.69$ | $(-32802.65, -32845.79)$ | |

Table 2.2

Parameter Estimates for The Translog SDF Model with Time-invariant Heterogeneity

| Parameter | Estimate | 95% Credible Interval | SIF |
|---|---|---|---|
| $a_1$ | 0.0742 | $(0.0549, 0.1022)$ | 59.9034 |
| $a_2$ | 0.5117 | $(0.5020, 0.5212)$ | 22.1721 |
| $a_3$ | 0.4013 | $(0.3526, 0.4393)$ | 60.8725 |
| $a_4$ | 0.0128 | $(-0.0054, 0.0354)$ | 54.8123 |
| $a_{11}$ | 0.0124 | $(0.0094, 0.0144)$ | 54.5859 |
| $a_{12}$ | $-0.006$ | $(-0.0093, -0.0026)$ | 29.3793 |
| $a_{13}$ | $-0.0099$ | $(-0.0143, -0.0053)$ | 44.1099 |
| $a_{14}$ | 0.0034 | $(0.0005, 0.0072)$ | 36.5554 |
| $a_{22}$ | 0.172 | $(0.1586, 0.1828)$ | 42.7098 |
| $a_{23}$ | $-0.1774$ | $(-0.1854, -0.1673)$ | 45.4679 |
| $a_{24}$ | 0.0114 | $(0.0044, 0.0175)$ | 28.0125 |
| $a_{33}$ | 0.1833 | $(0.1745, 0.1935)$ | 52.1131 |
| $a_{34}$ | 0.004 | $(-0.0006, 0.0110)$ | 30.1738 |
| $a_{44}$ | $-0.0188$ | $(-0.0314, -0.0112)$ | 41.2825 |
| $b_1$ | 0.0207 | $(-0.0197, 0.0538)$ | 55.2127 |
| $b_2$ | $-0.9062$ | $(-0.9675, -0.8418)$ | 59.9503 |
| $b_3$ | $-0.0377$ | $(-0.0671, -0.0131)$ | 55.4124 |
| $b_{11}$ | $-0.0547$ | $(-0.0932, -0.0249)$ | 44.0046 |
| $b_{12}$ | 0.0677 | $(0.0453, 0.0956)$ | 41.0148 |
| $b_{13}$ | $-0.0028$ | $(-0.0117, 0.0082)$ | 8.2088 |
| $b_{22}$ | $-0.0526$ | $(-0.0679, -0.0358)$ | 15.6079 |
| $b_{23}$ | $-0.0103$ | $(-0.0213, -0.0006)$ | 25.1706 |
| $b_{33}$ | $-0.001$ | $(-0.0117, 0.0094)$ | 29.5937 |
| $g_{11}$ | 0.0072 | $(0.0034, 0.0116)$ | 15.1585 |
| $g_{12}$ | $-0.0369$ | $(-0.0517, -0.0235)$ | 34.2838 |
| $g_{13}$ | 0.0097 | $(-0.0090, 0.0298)$ | 52.0561 |
| $g_{14}$ | 0.0199 | $(0.0084, 0.0343)$ | 34.9263 |
| $g_{21}$ | 0.0018 | $(-0.0055, 0.0106)$ | 49.6662 |
| $g_{22}$ | 0.0403 | $(0.0315, 0.0498)$ | 12.3706 |
| $g_{23}$ | $-0.0267$ | $(-0.0341, -0.0195)$ | 33.0743 |
| $g_{24}$ | $-0.0155$ | $(-0.0241, -0.0072)$ | 21.2348 |
| $g_{31}$ | $-0.0074$ | $(-0.0122, -0.0028)$ | 44.5503 |
| $g_{32}$ | $-0.0005$ | $(-0.0065, 0.0063)$ | 26.4733 |
| $g_{33}$ | 0.0057 | $(-0.0037, 0.0121)$ | 43.0810 |
| $g_{34}$ | 0.0022 | $(-0.0021, 0.0073)$ | 13.3574 |
| $\delta_1$ | $-0.0005$ | $(-0.0009, -0.0002)$ | 5.2898 |
| $\delta_2$ | 0.0009 | $(-0.0001, 0.0019)$ | 20.8112 |
| $\delta_3$ | $-0.0021$ | $(-0.0029, -0.0014)$ | 19.2264 |
| $\delta_4$ | 0.0018 | $(0.0008, 0.0028)$ | 26.1208 |
| $\delta_\tau$ | $-0.0007$ | $(-0.0012, -0.0003)$ | 23.1168 |
| $\delta_{\tau\tau}$ | 0.0002 | $(0.0000, 0.0004)$ | 11.9861 |
| $\rho_1$ | $-0.0021$ | $(-0.0038, -0.0002)$ | 20.3996 |
| $\rho_2$ | 0.0018 | $(0.0000, 0.0038)$ | 41.9750 |
| $\rho_3$ | $-0.0003$ | $(-0.0013, 0.0006)$ | 18.1842 |
| DIC | $-20715.33$ | $(-20676.63, -20763.89)$ | |

Table 2.3

Parameter Estimates for The Standard Translog SDF Model

| Parameter | Estimate | 95% Credible Interval | SIF |
|---|---|---|---|
| $a_1$ | $0.0537$ | $(0.0504, 0.0571)$ | $5.5117$ |
| $a_2$ | $0.5123$ | $(0.5053, 0.5197)$ | $3.5834$ |
| $a_3$ | $0.4385$ | $(0.4333, 0.4436)$ | $5.6730$ |
| $a_4$ | $-0.0046$ | $(-0.0101, 0.0011)$ | $4.9207$ |
| $a_{11}$ | $0.0162$ | $(0.0148, 0.0176)$ | $5.7111$ |
| $a_{12}$ | $-0.0060$ | $(-0.0085, -0.0033)$ | $9.5756$ |
| $a_{13}$ | $-0.0065$ | $(-0.0089, -0.0042)$ | $12.7311$ |
| $a_{14}$ | $-0.0037$ | $(-0.0059, -0.0017)$ | $7.2443$ |
| $a_{22}$ | $0.1807$ | $(0.1705, 0.1903)$ | $10.2052$ |
| $a_{23}$ | $-0.1621$ | $(-0.1684, -0.1556)$ | $13.7479$ |
| $a_{24}$ | $-0.0126$ | $(-0.0184, -0.0070)$ | $4.4540$ |
| $a_{33}$ | $0.1199$ | $(0.1178, 0.1220)$ | $10.5717$ |
| $a_{34}$ | $0.0487$ | $(0.0434, 0.0540)$ | $12.8721$ |
| $a_{44}$ | $-0.0323$ | $(-0.0394, -0.0258)$ | $10.2540$ |
| $b_1$ | $-0.0114$ | $(-0.0236, 0.0004)$ | $5.1877$ |
| $b_2$ | $-0.9802$ | $(-0.9910, -0.9695)$ | $4.5915$ |
| $b_3$ | $0.0004$ | $(-0.0075, 0.0083)$ | $6.0464$ |
| $b_{11}$ | $-0.0506$ | $(-0.0736, -0.0284)$ | $8.9720$ |
| $b_{12}$ | $0.0720$ | $(0.0555, 0.0892)$ | $8.8453$ |
| $b_{13}$ | $-0.0149$ | $(-0.0272, -0.0026)$ | $5.8968$ |
| $b_{22}$ | $-0.0519$ | $(-0.0674, -0.0366)$ | $6.8625$ |
| $b_{23}$ | $-0.0296$ | $(-0.0393, -0.0198)$ | $7.6843$ |
| $b_{33}$ | $0.0370$ | $(0.0289, 0.0451)$ | $4.7245$ |
| $g_{11}$ | $0.0103$ | $(0.0058, 0.0148)$ | $6.5304$ |
| $g_{12}$ | $-0.0419$ | $(-0.0559, -0.0281)$ | $9.4902$ |
| $g_{13}$ | $0.0025$ | $(-0.0074, 0.0120)$ | $13.3573$ |
| $g_{14}$ | $0.0291$ | $(0.0189, 0.0394)$ | $7.7076$ |
| $g_{21}$ | $-0.0059$ | $(-0.0095, -0.0022)$ | $7.9433$ |
| $g_{22}$ | $0.0541$ | $(0.0433, 0.0647)$ | $7.5070$ |
| $g_{23}$ | $-0.0294$ | $(-0.0358, -0.0228)$ | $9.3525$ |
| $g_{24}$ | $-0.0188$ | $(-0.0270, -0.0108)$ | $5.6318$ |
| $g_{31}$ | $-0.0083$ | $(-0.0113, -0.0055)$ | $4.3241$ |
| $g_{32}$ | $0.0044$ | $(-0.0030, 0.0119)$ | $8.4810$ |
| $g_{33}$ | $-0.0089$ | $(-0.0158, -0.0022)$ | $16.3715$ |
| $g_{34}$ | $0.0128$ | $(0.0073, 0.0182)$ | $7.9757$ |
| $\delta_1$ | $0.0008$ | $(0.0004, 0.0012)$ | $5.0648$ |
| $\delta_2$ | $-0.0014$ | $(-0.0023, -0.0005)$ | $4.8814$ |
| $\delta_3$ | $-0.0013$ | $(-0.0021, -0.0005)$ | $7.8134$ |
| $\delta_4$ | $0.0019$ | $(0.0010, 0.0027)$ | $7.6269$ |
| $\delta_\tau$ | $-0.0002$ | $(-0.0008, 0.0004)$ | $3.2046$ |
| $\delta_{\tau\tau}$ | $0.0000$ | $(-0.0002, 0.0002)$ | $3.0349$ |
| $\rho_1$ | $-0.0058$ | $(-0.0076, -0.0041)$ | $5.8436$ |
| $\rho_2$ | $0.0033$ | $(0.0021, 0.0046)$ | $5.8099$ |
| $\rho_3$ | $0.0017$ | $(0.0007, 0.0026)$ | $7.2146$ |
| DIC | $-17533.76$ | $(-17502.12, -17578.32)$ | |

Table 3

Mean Difference in Producitivity Growth

| | A. Between the Standard Translog SDF Model and the SDF Model with Time-varying Heterogeneity | | B. Between the SDF Model with Time-invariant Heterogeneity and the SDF Model with Time-varying Heterogeneity | |
|---|---|---|---|---|
| Year | Estimate | 95% Credible Interval | Estimate | 95% Credible Interval |
| 2005 | 0.0006 | $(-0.0033, 0.0045)$ | $-0.0051$ | $(-0.0105, -0.0005)$ |
| 2006 | $-0.0041$ | $(-0.0079, -0.0002)$ | $-0.0049$ | $(-0.0089, -0.0011)$ |
| 2007 | 0.0056 | $(0.0017, 0.0095)$ | $-0.0008$ | $(-0.0051, 0.0031)$ |
| 2008 | 0.0017 | $(-0.0022, 0.0055)$ | $-0.0041$ | $(-0.0088, -0.0001)$ |
| 2009 | 0.0022 | $(-0.0015, 0.0058)$ | $-0.0044$ | $(-0.0079, -0.0011)$ |
| 2010 | 0.0037 | $(0.0001, 0.0074)$ | $-0.0043$ | $(-0.0090, 0.0000)$ |
| 2011 | 0.0011 | $(-0.0024, 0.0046)$ | $-0.0031$ | $(-0.0069, 0.0004)$ |
| 2012 | 0.0050 | $(0.0014, 0.0084)$ | $-0.0023$ | $(-0.0052, 0.0005)$ |
| 2013 | $-0.0036$ | $(-0.0069, -0.0003)$ | $-0.0022$ | $(-0.0048, 0.0003)$ |
| Average | 0.0014 | $(-0.0023, 0.0050)$ | $-0.0035$ | $(-0.0075, 0.0002)$ |

Table 4

Technical Change

| Year | Estimate | 95% Credible Interval |
|---|---|---|
| 2004 | 0.0017 | $(0.0012, 0.0023)$ |
| 2005 | 0.0021 | $(0.0014, 0.0028)$ |
| 2006 | 0.0017 | $(0.0009, 0.0025)$ |
| 2007 | 0.0017 | $(0.0007, 0.0028)$ |
| 2008 | 0.0016 | $(0.0004, 0.0028)$ |
| 2009 | 0.0013 | $(-0.0002, 0.0027)$ |
| 2010 | 0.0012 | $(-0.0005, 0.0029)$ |
| 2011 | 0.0012 | $(-0.0008, 0.0032)$ |
| 2012 | 0.0012 | $(-0.0010, 0.0035)$ |
| 2013 | 0.0013 | $(-0.0012, 0.0038)$ |
| Average | 0.0016 | $(0.0003, 0.0028)$ |

Table 5.1

Technical Efficiency

| Year | Estimate | 95% Credible Interval |
|---|---|---|
| 2004 | 0.9786 | $(0.9767, 0.9803)$ |
| 2005 | 0.9777 | $(0.9755, 0.9798)$ |
| 2006 | 0.9798 | $(0.9778, 0.9817)$ |
| 2007 | 0.9753 | $(0.9731, 0.9776)$ |
| 2008 | 0.9713 | $(0.9691, 0.9735)$ |
| 2009 | 0.9744 | $(0.9724, 0.9764)$ |
| 2010 | 0.9723 | $(0.9701, 0.9745)$ |
| 2011 | 0.9714 | $(0.9691, 0.9739)$ |
| 2012 | 0.9686 | $(0.9661, 0.9712)$ |
| 2013 | 0.9695 | $(0.9666, 0.9723)$ |
| Average | 0.9744 | $(0.9729, 0.9760)$ |

Table 5.2

Distribution of Technical Efficiency Across BHCs

| Year | Minimum | Maximum | Standard Deviation | 5% percentile | 95% percentile |
|---|---|---|---|---|---|
| 2004 | 0.5381 | 0.9926 | 0.0316 | 0.9475 | 0.9908 |
| 2005 | 0.6541 | 0.9909 | 0.0240 | 0.9591 | 0.9879 |
| 2006 | 0.6896 | 0.9947 | 0.0198 | 0.9589 | 0.9902 |
| 2007 | 0.7495 | 0.9973 | 0.0233 | 0.9423 | 0.9917 |
| 2008 | 0.7948 | 0.9970 | 0.0292 | 0.9244 | 0.9937 |
| 2009 | 0.7430 | 0.9975 | 0.0301 | 0.9221 | 0.9939 |
| 2010 | 0.6467 | 0.9964 | 0.0350 | 0.9073 | 0.9932 |
| 2011 | 0.6184 | 0.9970 | 0.0393 | 0.9153 | 0.9942 |
| 2012 | 0.6052 | 0.9977 | 0.0430 | 0.9000 | 0.9939 |
| 2013 | 0.6472 | 0.9979 | 0.0401 | 0.9140 | 0.9949 |

Table 6.1

Returns To Scale

| Year | Estimate | 95% Credible Interval |
| --- | --- | --- |
| 2004 | 1.0034 | (1.0012, 1.0058) |
| 2005 | 1.0069 | (1.0045, 1.0093) |
| 2006 | 1.0054 | (1.0031, 1.0080) |
| 2007 | 1.0074 | (1.0048, 1.0102) |
| 2008 | 1.0078 | (1.0048, 1.0109) |
| 2009 | 1.0042 | (1.0008, 1.0077) |
| 2010 | 1.0023 | (0.9983, 1.0063) |
| 2011 | 1.0009 | (0.9965, 1.0055) |
| 2012 | 1.0014 | (0.9966, 1.0065) |
| 2013 | 1.0016 | (0.9964, 1.0071) |
| Average | 1.0044 | (1.0013, 1.0076) |

Table 6.2

Returns To Scale at Individual BHC Level

| Year | Increasing Returns to Scale | Constant Returns to Scale | Decreasing Returns to Scale |
|---|---|---|---|
| 2004 | 55.22% | 25.37% | 19.40% |
| 2005 | 67.46% | 19.10% | 13.43% |
| 2006 | 59.40% | 24.18% | 16.42% |
| 2007 | 64.71% | 23.53% | 11.76% |
| 2008 | 65.85% | 22.54% | 11.62% |
| 2009 | 52.75% | 30.40% | 16.85% |
| 2010 | 43.19% | 37.35% | 19.46% |
| 2011 | 37.45% | 41.98% | 20.58% |
| 2012 | 38.43% | 40.17% | 21.40% |
| 2013 | 38.99% | 43.12% | 17.89% |
| Average | 53.78% | 29.59% | 16.63% |

Table 7

Productivity Growth and Its Decomposition

| Year | Productivity growth | | Efficiency Change | | Technical Change | | Scale Effect | |
|---|---|---|---|---|---|---|---|---|
| | Estimate | 95% CI | Estimate | 95% CI | Estimate | 95% CI | Estimate | 95% CI |
| 2005 | 0.0017 | (-0.0005, 0.0039) | -0.0009 | (-0.0030, 0.0011) | 0.0021 | (0.0014, 0.0028) | 0.0005 | (0.0004, 0.0007) |
| 2006 | 0.0049 | (0.0028, 0.0069) | 0.0021 | (0.0003, 0.0039) | 0.0017 | (0.0009, 0.0025) | 0.0011 | (0.0008, 0.0013) |
| 2007 | -0.0029 | (-0.0054, -0.0005) | -0.0052 | (-0.0076, -0.0030) | 0.0017 | (0.0007, 0.0028) | 0.0006 | (0.0005, 0.0007) |
| 2008 | -0.0018 | (-0.0040, 0.0005) | -0.0042 | (-0.0062, -0.0024) | 0.0016 | (0.0004, 0.0028) | 0.0008 | (0.0006, 0.0010) |
| 2009 | 0.0027 | (0.0005, 0.0049) | 0.0016 | (-0.0001, 0.0032) | 0.0013 | (-0.0002, 0.0027) | -0.0002 | (-0.0003, -0.0001) |
| 2010 | -0.0032 | (-0.0056, -0.0009) | -0.0042 | (-0.0058, -0.0025) | 0.0012 | (-0.0005, 0.0029) | -0.0003 | (-0.0004, -0.0002) |
| 2011 | -0.0004 | (-0.0028, 0.0020) | -0.0016 | (-0.0033, 0.0000) | 0.0012 | (-0.0008, 0.0032) | 0.0001 | (0.0000, 0.0001) |
| 2012 | -0.0024 | (-0.0048, 0.0001) | -0.0038 | (-0.0055, -0.0020) | 0.0012 | (-0.0010, 0.0035) | 0.0002 | (0.0001, 0.0004) |
| 2013 | 0.0010 | (-0.0014, 0.0035) | -0.0005 | (-0.0022, 0.0013) | 0.0013 | (-0.0012, 0.0038) | 0.0002 | (0.0001, 0.0003) |
| Average | 0.0002 | (-0.0012, 0.0015) | -0.0018 | (-0.0021, -0.0014) | 0.0016 | (0.0002, 0.0029) | 0.0004 | (0.0003, 0.0005) |